Universiteit Utrecht

Master Kinder- en Jeugdpsychologie

Betrouwbaarheid en Validiteit van een Productieve en Receptieve Woordenschattest voor Jonge Tweedetaalverwervers.

Stephanie J. Hoeber

3519422

27 juni 2014

Begeleiders:

Dr. Dorian de Haan, universitair docent Kind- en Jeugdpsychologie, Universiteit van Utrecht

Dr. Herman van Boxtel, universitair hoofddocent Kind- en Jeugdpsychologie, Universiteit van Utrecht

(tweede beoordelaar)

Abstract

This research investigates the reliability and validity of a productive and a receptive vocabulary test. These tests aim to measure the effect of a new approach in vocabulary teaching for young second language learners called 'Total Physical Response' (TPR). Forty-seven four-year-old second language learners took part in this research. The passive vocabulary task of the 'Taaltoets Alle Kinderen' was conducted to investigate the construct validity. Both the productive vocabulary test (r =.78) and the receptive vocabulary test (r= .71) correlated strongly with the scores on the passive vocabulary task and were very reliable, respectively .86 and .88. The tests also correlated strongly with each other (r= .78). The tests are reliable and valid measures of the vocabulary among young second language learners. The productive vocabulary test is very difficult and could cause stress on children with a limited vocabulary.

Keywords: Total Physical Response, vocabulary test, second language learners, reliability, construct validity.

Abstract

Huidig onderzoek stelt de betrouwbaarheid en validiteit van een productieve en een receptieve woordenschattest vast. Deze testen zullen gebruikt worden om het effect van een nieuwe aanpak van het woordenschatonderwijs voor tweedetaalverwervers, genaamd 'Total Physcial Response' (TPR), te onderzoeken. Aan het onderzoek hebben 47 vierjarige tweedetaalverwervers deelgenomen. De passieve woordenschattaak van de 'Taaltoets Alle Kinderen' is afgenomen om de begripsvaliditeit te onderzoeken. Zowel de scores op de productieve woordenschattest (r =.78) als de scores op de receptieve woordenschattest (r= .71) correleren sterk met de scores op de passieve woordenschattaak. Beide testen hebben ook een goede betrouwbaarheid, respectievelijk .86 en .88. De testen correleren ook sterk met elkaar (r= .78). De testen zijn betrouwbaar en valide in het meten van de woordenschat van jonge tweedetaalverwervers. De afname van de productieve woordenschattest kan stressvol zijn voor kinderen met een beperkte woordenschat, omdat deze test erg moeilijk is.

Trefwoorden: Total Physical Response, woordenschattest, tweedetaalverwervers, betrouwbaarheid, begripsvaliditeit.

Betrouwbaarheid en Validiteit van een Productieve en Receptieve Woordenschattest voor Jonge Tweedetaalverwervers.

Jonge tweedetaalverwervers zijn kinderen die thuis weinig of geen Nederlands spreken, waardoor zij bij binnenkomst op de basisschool nog aan het begin staan van het leren van de Nederlandse taal. Deze jonge tweedetaalverwervers hebben al bij de start van hun loopbaan in het basisonderwijs een gemiddelde taalachterstand van twee jaar (Driessen & Doesborgh, 2003; Tesser & Iedema, 2001). Kinderen van Turkse of Marokkaanse afkomst hebben de grootste taalachterstand (Hartgers, 2012; Tesser & Iedema, 2001). Zo loopt bijvoorbeeld de woordenschat van kinderen van Turkse of Marokkaanse afkomst ver achter op de woordenschat van kinderen van Nederlandste afkomst (Verhoeven & Vermeer, 1992). De belangrijkste oorzaken zijn het lagere opleidingsniveau van de ouders, de taalvaardigheid van de ouders en de taal die thuis wordt gesproken (Gijsberts, Huijnk & Dagevos, 2012; Hartgers, 2012; Tesser & Iedema, 2001). In 2011 werd in meer dan de helft van de Turkse en een derde van de Marokkaanse gezinnen thuis geen Nederlands gesproken (Hartgers, 2012).

Het onderwijs zal extra aandacht moeten bieden aan de Nederlandse taalverwerving van deze jonge tweedetaalverwervers. De focus zal moeten liggen op het vergroten van de woordenschat, omdat voldoende woordkennis bepalend is voor de taalvaardigheid van een kind en het latere schoolsucces (Verhallen & Verhallen, geciteerd in Bacchini, Boland, Hulsbeek, Pot & Smits, 2005). Op dit moment lijkt het onderwijs nog onvoldoende in staat de achterstanden in de woordenschat weg te werken (Driessen & Doesborgh, 2003). Door de beperkte Nederlandse woordenschat van deze kinderen begrijpen zij niet wat de leerkracht zegt, waardoor zij niet goed mee kunnen doen in de klas (Verhoeven, van Leeuwe & Vermeer, 2011). De kinderen die hun schooltijd wel beginnen met een omvangrijke woordenschat kunnen de leerkracht gemakkelijk volgen, begrijpen de voorleesverhalen en opdrachten beter, kunnen in de kring goed meedoen en leren dan ook snel de betekenis van allerlei nieuwe woorden (Bacchini et al., 2005). “Het is immers zo, dat hoe meer woorden je al kent, des te gemakkelijker je weer nieuwe woorden leert” (Bacchini

et al, 2005, p. 6). De beperkte woordenschat zorgt er tevens voor dat deze kinderen minder van het onderwijs kunnen profiteren (Bacchini, 2012). De kinderen lopen hierdoor ook een leerachterstand op. Deze leerachterstand is tijdens het voorgezet onderwijs en het beroepsonderwijs nog steeds aanwezig (Hartgers, 2012).

Aangezien jonge tweedetaalverwervers door hun beperkte woordenschat minder van het onderwijs profiteren, is het belangrijk om deze achterstanden al voor de start van het onderwijs aan te pakken. Een VVE-programma is een programma dat zich richt op het voorkomen en bestrijden van onderwijsachterstanden in de voor- en vroegschoolse periode, waaronder de achterstanden in de woordenschat (Meij, Mutsaers & Pennings, 2009). Het 'Effective Provision of Pre-School Education Project' (EPPE) onderzocht in Groot-Brittannië de effecten van VVE-programma's bij kinderen tussen de drie en zeven jaar. VVE-programma's van hoge kwaliteit bleken een positieve invloed te hebben op de intellectuele en sociale ontwikkeling van kinderen (Sylva, Melhuish, Sammons, Siraj-Blatchford & Taggert, 2004). Ook werd een relatie gevonden tussen het volgen van een VVE-programma en hogere scores op lees- en rekentesten op latere leeftijd (Sylva et. al, 2004). VVE-programma's zouden een geschikt middel kunnen zijn om de achterstanden in de woordenschat van deze tweedetaalverwervers op jonge leeftijd weg te werken.

VVE-programma's

Sinds de komst van de regeling 'Voor- en Vroegschoolse Educatie' in 2000 en de wet 'Ontwikkelingskansen door Kwaliteit en Educatie' in 2010 van het ministerie van Onderwijs, Cultuur en Wetenschap zijn VVE-programma's een belangrijk onderdeel van het onderwijsachterstandenbeleid. Één van de doelstellingen van het VVE-beleid is dat in de komende jaren alle kinderen in achterstandsituaties worden bereikt en dat hen optimale ontwikkelingskansen worden aangeboden (Ledoux & Veen, 2009). Als gevolg van dit beleid ontstond er een ruime keuze uit VVE-programma's. In de VVE-keuzegids (Dekker & Gelauf-Hanzon, 2000) worden 51 verschillende programma's omschreven. Een voorbeeld van een VVE-programma is 'Ik en Ko'. Dit programma heeft als hoofddoel het vergroten van de taalvaardigheid. Net als vele andere VVE-programma's is

‘Ik en Ko’ thematisch opgebouwd met thema’s als “Brr, wat koud”, “Bij mij thuis”, “Eet smakelijk” en “Circus”. Een thema duurt drie tot vier weken, waarin verschillende activiteiten zijn uitgewerkt, zoals discussies, ontdekactiviteiten, fantasiespel en knutselen. Per activiteit is aangegeven hoe de leerkracht de activiteit kan voorbereiden en welke woorden de leerkracht aan zou kunnen bieden. De thema’s zijn flexibel inzetbaar en de leerkracht kan zelf bepalen wanneer een thema aan bod komt (Pennings, 2006).

Er is in Nederland echter maar weinig onderzoek gedaan naar de uitvoering en de effectiviteit van VVE-programma’s (Driessen, 2012; Meij et al., 2009; Pot, 2006). Zo is er ook nog geen onderzoek verricht naar de effectiviteit van ‘Ik en Ko’ (Nap-Kolhoff, van Schilt-Mol, Sontag, van Steensel & Vallen, 2008; Pennings, 2006). Uit het wel uitgevoerde onderzoek blijkt dat VVE-programma’s in Nederland onvoldoende effectief zijn in het vroegtijdig wegwerken van onder andere achterstanden in de woordenschat (de Haan, Leseman, Elbers, 2011; Helms-Lorenz & de Jong-Heeringa, 2006; de Jong-Heeringa & Bosker, 2006; Veen, Fukkink & Roeleveld, 2006; Veen, Roeleveld & Leseman, 2000). Dit komt mogelijk doordat het taalaanbod in de activiteiten, de voorleestekstjes en liedjes binnen deze programma’s niet te begrijpen zijn voor jonge tweedetaalverwervers (Pot, 2006). De werkwijze van VVE-programma’s sluit onvoldoende aan bij het taalniveau van deze kinderen, die vaak minder dan 1000 Nederlandse woorden receptief kennen (Bacchini et al., 2005).

Total Physical Response (TPR)

‘Total Physical Response’ (Asher, 1988) wordt wereldwijd ingezet voor het leren van een tweede of vreemde taal aan beginnende taalleerders van alle leeftijden. Bij deze aanpak spelen de begrippen interactie en begrijpelijke input een grote rol (Asher, 1988). TPR is gebaseerd op het proces van de eerste taalverwerving. Kinderen beginnen niet meteen met het spreken van hun moedertaal. Daar gaan maanden aan vooraf, waarbij kinderen zonder taal te produceren toch interacteren met hun ouders. Vaak geven ouders opdrachten, zoals ‘neem een hap’, ‘hou mijn hand

vast' of 'kom maar hier'. Het kind reageert vervolgens met een fysieke actie. Deze taal-lichaam conversaties vormen de basis van TPR (Asher, 1988).

De beginnende leerling krijgt met behulp van attributen en visuele ondersteuning korte opdrachten (Pot, 2012). Leerlingen worden op hun eigen niveau aangesproken en mogen op hun eigen niveau reageren, eerst non-verbaal en daarna eventueel ook verbaal (Pot, 2012). De leerlingen leren door middel van deze doe-opdrachten via de receptieve weg. Dit betekent dat leerlingen eerst de betekenis van woorden leren voordat zij deze woorden gaan produceren. Uit kleinschalig onderzoek van Pot (2012) blijkt TPR een veelbelovend programma voor het vergroten van de woordenschat van beginnende taalleerders. Er is echter grootschalig onderzoek nodig om te kunnen bepalen of TPR de woordenschat vergroot van jonge tweedetaalverwervers. In de nabije toekomst zal de effectiviteit van de methodiek TPR binnen het thema 'Circus' van 'Ik en Ko' worden onderzocht in quasi-experimenteel onderzoek met een experimentele conditie (reguliere aanpak met TPR) en een controle conditie (reguliere aanpak zonder TPR). 'Ik en Ko' is gekozen voor het onderzoek naar de effectiviteit van TPR, omdat door de thematische opbouw van 'Ik en Ko' het aanbod van doelwoorden over de verschillende condities gelijk kan worden gehouden. Hiernaast is voor 'Ik en Ko' gekozen, omdat al veel scholen werken met dit programma en inmiddels gewend zijn aan de werkwijze van 'Ik en Ko'. Vooraf aan het huidige onderzoek is een productieve woordenschattest en een receptieve woordenschattest ontwikkeld. De productieve woordenschattest meet het aantal woorden dat een kind kan produceren en de receptieve woordenschattest meet het aantal woorden dat een kind begrijpt. Deze testen zullen tijdens het effectiviteitsonderzoek naar TPR gebruikt worden om vast te stellen of kinderen meer woorden kunnen leren met behulp van TPR dan met behulp van de reguliere aanpak van 'Ik en Ko'. Om het effect van TPR op de productieve en receptieve woordenschat te kunnen onderzoeken, is het belangrijk om de psychometrische kwaliteit van de productieve en receptieve woordenschattesten vast te stellen. Deze thesis bevat het verslag van een onderzoek naar de betrouwbaarheid en validiteit van de productieve en de receptieve woordenschattesten.

Betrouwbaarheid en validiteit van de woordenschattesten

Het is belangrijk goed onderscheid te kunnen maken tussen de betrouwbaarheid en de validiteit van een test. Een test is betrouwbaar als het nauwkeurig en zuiver een bepaalde vaardigheid of een bepaald construct meet (Goodwin, 2008). Een test is valide wanneer het de vaardigheid of het construct meet die het beoogt te meten (Goodwin, 2008). Een test moet zowel betrouwbaar als valide zijn om een goede test te zijn. Daarnaast wordt de kwaliteit van een test bepaald door de kwaliteit van zijn afzonderlijke items.

Bij het vast stellen van de kwaliteit van de afzonderlijke items is het allereerst van belang dat alle items hetzelfde construct meten, namelijk de productieve of receptieve woordenschat. Dit valt af te leiden uit de item-restcorrelaties van de items. Dit is de correlatie tussen de scores op een item en de scores op de rest van de test (Field, 2009). Een item dat laag of zelfs negatief correleert met de rest van een test meet mogelijk een ander construct en zal moeten worden aangepast of verwijderd (Field, 2009). De item-restcorrelatie van een item geeft ook informatie over het discriminerend vermogen van dit item. Items met een goed discriminerend vermogen zijn in staat onderscheid te maken tussen kinderen met een hoge totaalscore en kinderen met een lage totaalscore (Dousma, Horsten & Brants, 1997). Een hoge positieve item-restcorrelatie van een item houdt in dat kinderen die goed hebben gepresteerd op de gehele test dit item vaker juist hebben beantwoord dan kinderen die slechter hebben gepresteerd op de gehele test. Ebel (1972) heeft normen vastgesteld voor de hoogte van de item-restcorrelaties. Deze normen worden weergegeven in tabel 1.

Tabel 1

Normen voor Item-restcorrelaties (naar Ebel, 1972)

Item-restcorrelatie	Itembeoordeling
0.40 en hoger	zeer goed
0.30 tot 0.39	goed, voor verbetering vatbaar
0.20 tot 0.29	twijfelgeval, moet verbeterd worden
0.19 en lager	slecht, verwijderen of herzien

Vervolgens is het informatief om de moeilijkheidsgraden van items te bekijken. Wanneer items zeer makkelijk zijn, zullen de meeste kinderen de items goed beantwoorden. Wanneer items zeer moeilijk zijn, zullen de meeste kinderen de items fout beantwoorden. Extreem gemakkelijke of moeilijke items maken om deze reden geen onderscheid tussen bijvoorbeeld kinderen met een beperkte woordenschat en kinderen met een uitgebreide woordenschat. De moeilijkheidsgraad van een item is de proportie kinderen die het goede antwoord heeft gegeven en wordt uitgedrukt met een p-waarde (Bender, 1994). De optimale p-waarde van een item is .50. Wanneer items een p-waarde van .50 hebben, wordt optimaal gediscrimineerd tussen kinderen met een lage totaalscore en kinderen met een hoge totaalscore en zullen de item-restcorrelaties van de items en de totale betrouwbaarheid van een test maximaal zijn (Feldt, 1993). Wanneer sprake is van een bepaalde gokkans, zoals bij een meerkeuzevraag, stijgt de optimale p-waarde naar .57 tot .67 (Feldt, 1993). Items met een p-waarden tussen de .27 en .79 hebben geen negatieve invloed op de betrouwbaarheid en kunnen behouden worden (Feldt, 1993).

De kwaliteit van een test wordt ook bepaald door de betrouwbaarheid van de test. Een instrument is betrouwbaar als het nauwkeurig en zuiver een bepaalde vaardigheid meet (Cohen & Swerdlik, 2010). Volgens de klassieke testtheorie bestaat een score uit een ware score op een bepaalde vaardigheid en uit meetfouten door toeval (Henson, 2001). Of er sprake is van meetfouten door toeval kan worden onderzocht door een meting te herhalen. Wanneer bij de herhaling van een

test veel overeenstemming is tussen de scores op de metingen, is er minder sprake van fouten op basis van toeval en heeft de test een hoge betrouwbaarheid (Cohen et al., 2010). Één van de meest gebruikte schattingen van de betrouwbaarheid van een test is het vast stellen van de interne consistentie van een test. Het is voor deze schatting namelijk niet nodig om een test meerdere keren af te nemen (Henson, 2001). De interne consistentie van een test is de mate waarin de items van de test met elkaar samenhangen (Sijtsma, 2009). De verschillende items van een test worden als hele korte sub-tests opgevat die elkaar opvolgen. Interne consistentie is geen directe maat van betrouwbaarheid, maar een theoretische schatting op basis van de klassieke testtheorie. Meer directe schattingen van de betrouwbaarheid zijn test-hertest-betrouwbaarheid en alternatieve vorm betrouwbaarheid, waarbij de consistentie in scores over tijd en consistentie in scores tussen testen wordt onderzocht (Henson, 2001).

Vervolgens is het belangrijk dat de productieve en receptieve woordenschattesten daadwerkelijk de woordenschat van jonge tweedetaalverwervers meten en niet een ander construct. Dit wordt de begripsvaliditeit van een instrument genoemd (Cohen et al., 2010). De begripsvaliditeit van een test wordt onder andere onderzocht door de scores op de test te vergelijken met de scores op een test waarvan al helemaal is vastgesteld dat het dezelfde vaardigheid betrouwbaar en valide meet (Cohen et al., 2010). Wanneer de scores op de testen hoog met elkaar correleren, kan geconcludeerd worden dat de testen waarschijnlijk hetzelfde construct meten.

Naast de betrouwbaarheid en de validiteit van de testen is ook de bruikbaarheid van de testen bij het effectiviteitsonderzoek naar TPR van belang. Wanneer de items van de testen te makkelijk zijn en de kinderen bij de voormeting de meeste items al goed beantwoorden, is het niet mogelijk om een progressie in de woordenschat van de kinderen te meten. Er kan dan niet meer vastgesteld worden of TPR een positieve invloed heeft op de woordenschat van kinderen. Wanneer de items juist te moeilijk zijn, kan dit stress bij de kinderen veroorzaken en de kinderen demotiveren om nogmaals mee te werken aan één van de testen. Ook zal de duur van de instrumenten invloed

hebben op hun bruikbaarheid. Kinderen van vier jaar kunnen gemiddeld slechts 10 minuten de aandacht richten op één taak (Sarid & Breznitz, 1997).

Om de kwaliteit van de productieve en receptieve woordenschattesten vast te stellen, is gekeken naar de kwaliteit van de afzonderlijke items aan de hand van de item-restcorrelaties en de moeilijkheidsgraden. Vervolgens zijn de betrouwbaarheid en de begripsvaliditeit van de testen onderzocht. Er is gekeken naar de intern consistentie van de testen en de mate waarin de scores op de productieve woordenschattest en de receptieve woordenschattest overeenkomen met de scores op een woordenschattest waarvan de betrouwbaarheid en validiteit al is vastgesteld. Ten slotte is gekeken of de testen bruikbaar zijn voor het effectiviteitsonderzoek naar TPR.

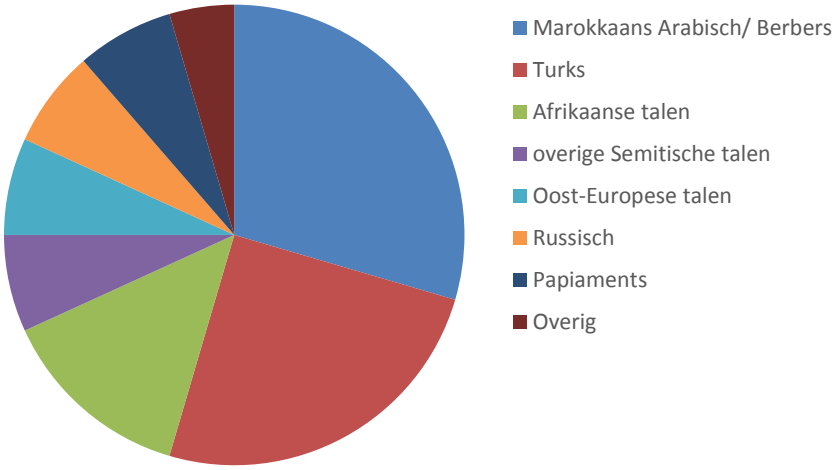
Methode

Participanten

Om de psychometrische kwaliteit en de bruikbaarheid van de productieve en de receptieve woordenschattesten te kunnen bepalen, is het nodig de testen af te nemen onder vierjarige tweedetaalverwervers, aangezien de effectiviteit van TPR ook onder deze doelgroep zal worden onderzocht. Er is besloten om tijdens het huidige onderzoek, in tegenstelling tot het effectiviteitsonderzoek naar TPR, naast kinderen met een beperkte woordenschat ook kinderen met een gemiddelde of uitgebreide woordenschat mee te nemen. Er is namelijk voldoende spreiding in testscores nodig om de betrouwbaarheid en de validiteit van een test te kunnen bepalen. Een beperkt bereik van scores heeft namelijk een negatief effect op de grootte van correlaties (Goodwin & Leech, 2006). Het effect van een beperkt bereik op de correlaties tussen de scores op de testen wordt geïllustreerd in de figuren A1 en A2 in bijlage A. De verbanden tussen de testen zijn veel minder duidelijk, wanneer alleen wordt gekeken naar de scores van de kinderen met een beperkte woordenschat.

Voor huidig onderzoek zijn basisscholen benaderd binnen Zuid-Holland. Bijlage B bevat de informatiebrief die naar de scholen is gestuurd. Een kind werd geselecteerd wanneer het vier jaar was en de leerkracht aangaf dat het kind thuis geen of weinig Nederlands spreekt. Van te voren

werd ook een toestemmingsformulier voor de ouders (zie bijlage C) naar de scholen gestuurd, zodat ouders konden aangeven of zij het eens waren met de deelname van hun kind. Geen van de ouders of de kinderen heeft afgezien van deelname. De productieve en de receptieve woordenschattesten zijn in totaal bij 47 vierjarige tweedetaalverwervers afgenomen. Drie kinderen zijn niet meegenomen in de analyses, omdat deze kinderen kort voor de afname van de testen les hebben gehad in het thema 'Circus' van 'Ik en Ko'. Dit was terug te zien in hun scores op de productieve en receptieve woordenschattesten, die gebaseerd zijn op dit thema van 'Ik en Ko'. Uiteindelijk zijn 44 kinderen meegenomen in de data-analyse. Er is bij elk kind gevraagd naar de precieze leeftijd, de sekse en de gesproken thuistaal. Deze groep bestond uit 26 jongens en 18 meisjes. De jongens waren gemiddeld 55 maanden oud (SD= 3.55) en de meisjes waren gemiddeld 54 maanden oud (SD= 3.46). In figuur 1 is te zien dat iets meer dan de helft van de kinderen thuis Turks of een taal uit Marokko spreekt. De andere helft van de kinderen spreken thuis veel verschillende talen, daarom zijn een groot deel van de overige talen ondergebracht in subgroepen.



Figuur 1. De taal die volgens de leerkracht bij het kind thuis wordt gesproken.

Instrumenten

Productieve woordenschattest met het thema 'Circus'. Bij deze test wordt gebruik gemaakt van een platenboek. Op elke bladzijde wordt een afbeelding weergegeven die het kind moet benoemen. Een voorbeeld van een afbeelding uit productieve test is te zien in figuur 2.

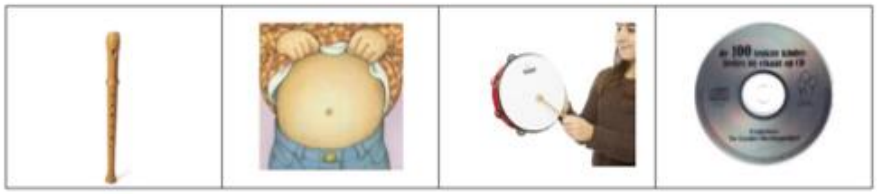
Dit is een.....

Figuur 2. Een afbeelding uit de productieve woordenschattest.

Wanneer het antwoord van een kind in de juiste richting zat maar het exacte doelwoord nog niet werd geproduceerd, werd doorgevraagd. Een antwoord werd alleen goed gerekend wanneer de doelwoorden werden geproduceerd. Voor kinderen onder de 5 jaar zijn medeklinkercombinaties (stempelen; clown) en lange woorden (goochelaar/ directeur) nog moeilijk uit te spreken (Gillis & Schaerlaekens, 2000). Hiermee is rekening gehouden bij het goed rekenen van antwoorden.

Receptieve woordenschattest met het thema 'Circus'. Ook bij de receptieve woordenschattest wordt gebruik gemaakt van een platenboek. Bij deze test wordt het kind per item gevraagd uit vier plaatjes te kiezen. De vier plaatjes bestaan uit één referent en drie afleiders. Er is bij elk item geprobeerd om van de drie afleiders er één in klank overeen te laten komen, één uit hetzelfde domein te kiezen en één te kiezen die geen overeenkomst heeft met de referent. Bij elk item is gezocht naar passende afleiders van eenzelfde taalniveau, gemeten aan de BAK- en WAK-

lijsten (Kuiken & Droge, 2010; Mulder, Timman & Verhallen, 2009). In figuur 3 wordt een item van de receptieve woordenschattest afgebeeld.



Figuur 3. Een item uit de receptieve woordenschattest: *"Een fluit. Waar zie je een fluit?".*

Bij deze test zijn de doelwoorden eerst 'los' aangeboden voordat de vraag werd gesteld. Bacchini (2012) stelt dat jonge tweedetaalverwervers nieuwe woorden moeilijk kunnen isoleren uit de stroom klanken binnen een zin. Losse woorden worden beter herkend. Om deze reden is besloten om de vragen op de volgende manier te stellen*: "Lopen. Wie is aan het lopen?".*

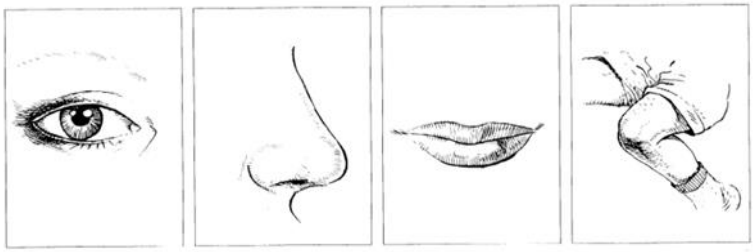
De productieve en receptieve woordenschattesten zijn beide gebaseerd op het thema 'Circus' van het VVE-programma 'Ik & Ko'. Voor beide testen zijn dezelfde 30 doelwoorden geselecteerd, waarvan 16 zelfstandig naamwoorden, 8 werkwoorden en 6 uit de categorie 'overige woorden'. Er zijn vooral doelwoorden gekozen die goed afbeeldbaar zijn en begrijpbaar zijn voor vierjarigen zonder veel kennis van de wereld of van de Nederlandse cultuur. Doordat de twee testen dezelfde doelwoorden testen, zijn de afbeeldingen van de twee woordenschattesten vergelijkbaar. Toch is geprobeerd om de afbeeldingen niet te veel op elkaar te laten lijken. Wanneer de afbeeldingen van de testen te veel op elkaar lijken, kunnen de testen elkaar beïnvloeden.

Er is bij deze testen geprobeerd om de instructies zo makkelijk mogelijk te houden, zodat zwakke resultaten niet in termen van niet verstaan of niet begrijpen kunnen worden uitgelegd. De meeste woorden van de instructies behoren tot de eerste 200 woorden die kinderen kennen (Bacchini, 2012). De vraagwoorden *wie, wat, waar* behoren echter tot de eerste 1000 woorden (Bacchini, 2012).

Met behulp van de 'Basiswoordenlijst Amsterdamse Kinderen' (BAK) voor groep 1 en 2 (Mulder et al., 2009) en van de 'Woordenlijst Amsterdamse Kinderen' (WAK) voor groep 3 tot en met 8 (Kuiken et al., 2010) is geprobeerd de items van makkelijk naar moeilijk te rangschikken. De BAK en WAK lijsten geven aan welke woorden de kinderen per groep moeten kennen om goed in het onderwijs mee te kunnen doen. Er is echter nooit onderzoek gedaan naar de moeilijkheid van de woorden die zijn opgenomen in deze lijsten. Het is hierdoor niet met zekerheid te stellen of de items daadwerkelijk van makkelijk naar moeilijk zijn gerangschikt. Tijdens het huidige onderzoek zijn moeilijkheidsgraden berekend, waardoor gecontroleerd kan worden of de items van makkelijk naar moeilijk zijn gerangschikt.

Passieve woordenschattaak van de 'Taaltoets Alle Kinderen' (TAK). De passieve woordenschattaak van de 'Taaltoets Alle Kinderen' (TAK) van Cito is afgenomen om de begripsvaliditeit van de productieve en receptieve woordenschattesten vast te stellen. Deze test meet de receptieve woordenschat van kinderen tussen de vier en negen jaar met Nederlands als moedertaal of als tweede taal. De betrouwbaarheid en validiteit van de TAK is door COTAN beoordeeld als goed. Scores op de TAK komen overeen met beoordelingen van leraren en geanalyseerde gespreksopnames (Verhoeven & Vermeer, 2006). Hiernaast bestaat er een relatie tussen de scores op de TAK in groep 1 en de taalvaardigheid van het kind in groep 4 (Verhoeven et al.,2006). Ook kunnen scores van de TAK in groep 1 de score op de Cito-toets in groep 8, de Eindtoets Basisonderwijs, voorspellen (Verhoeven et al., 2006).

Bij de passieve woordenschattaak van de TAK wordt ook gebruik gemaakt van een platenboek. In totaal is bij elk van 96 woorden een viertal tekeningen gemaakt, waarvan één referent en drie afleiders (Verhoeven et al., 2006). Ook bij deze taak wordt een kind gevraagd het juiste plaatje aan te wijzen. Een voorbeelditem van deze taak wordt afgebeeld in figuur 4.



Figuur 4. Item uit de passieve woordenschattaak van de 'Taaltoets Alle Kinderen' (TAK) van Cito: "*Waar zie je een neus?*".

Procedure

De testen zijn op maandag- en dinsdagochtend afgenomen, wanneer de kinderen nog de meeste energie hebben. Voor de afname van de testen zijn de kinderen één voor één uit de klas gehaald en meegenomen naar een rustige ruimte. De testleider ging altijd tegenover het kind zitten, wegens de vormgeving van de draaiboeken van de drie testen. De afname van de testen duurde bij elkaar ongeveer 15 minuten. De drie verschillende testen werden niet in één keer afgenomen, aangezien kinderen van vier jaar gemiddeld slechts 10 minuten de aandacht kunnen richten op één taak (Sarid et al., 1997). De kinderen zijn om deze reden twee keer uit de klas gehaald. De eerste keer werd de passieve woordenschattaak van de TAK afgenomen. Tijdens de tweede ronde zijn de woordenschattesten met het thema 'Circus' afgenomen. Er is besloten om eerst de productieve woordenschattest af te nemen en vervolgens pas de receptieve woordenschattest, omdat deze testen dezelfde doelwoorden testen. Wanneer de receptieve woordenschattest vooraf aan de productieve test wordt afgenomen, worden alle doelwoorden al voorgezegd door de testleider. Dit zou de resultaten van de productieve woordenschattest beïnvloeden. Deze beide testen startten met twee voorbeelditems. Alleen tijdens deze twee items kreeg het kind feedback op zijn prestatie. Tijdens de rest van de testen werd dit niet meer gedaan. Deze voorbeelditems zijn niet meegenomen in de data-analyse.

Data-analyse

Allereerst zijn de item-restcorrelaties van de afzonderlijke items bekeken om te bepalen of de items goed met elkaar samenhangen. Volgens de normen van Ebel (1972) moeten items met lagere item-restcorrelaties dan .30 aangepast worden en items met lagere item-restcorrelaties dan .20 verwijderd of herzien (zie tabel 1). De items met lagere item-restcorrelatie dan .20 zijn tijdelijk verwijderd uit de testen om de invloed van deze items op de betrouwbaarheid en de validiteit van de testen vast te stellen.

Ook zijn de moeilijkheidsgraden van de items bekeken. Deze geven een goede indicatie van de bruikbaarheid van de testen. Zoals eerder is besproken hebben items met lagere moeilijkheidsgraden dan .27 of hogere moeilijkheidsgraden dan .79 een negatieve invloed op de betrouwbaarheid (Feldt, 1993). Bij een steekproefgrootte van 44 participanten bevatten de moeilijkheidsgraden echter nog een meetfout van ongeveer .07 (Dousma et al., 1997). Er is om deze reden gekeken of items met lagere moeilijkheidsgraden dan .27 of hogere moeilijkheidsgraden dan .79 daadwerkelijk een negatieve invloed hebben op de betrouwbaarheid.

Voor de items van de receptieve woordenschattest zijn gecorrigeerde moeilijkheidsgraden berekend, waarbij gecorrigeerd is voor de gokkans waarvan sprake is bij meerkeuzevragen (Dousma et al., 1997). Ook als een kind niet het antwoord op een item weet, is er een bepaalde kans dat het kind toch het juiste antwoord geeft. Dit leidt tot hogere moeilijkheidsgraden en daar moet voor worden gecorrigeerd.

De betrouwbaarheid van de testen is onderzocht aan de hand van Guttman's lambda2 (Guttman, 1945). Over het algemeen wordt Cronbachs (1951) alfa gebruikt om de betrouwbaarheid van een test te schatten. Guttman's lambda2 geeft echter een betere schatting van de betrouwbaarheid, omdat Guttman's lambda2 in mindere mate de betrouwbaarheid van een test systematisch onderschat (Evers, Lucassen, Meijers & Sijtsma, 2010).

Ten slotte is de begripsvaliditeit van de productieve en receptieve woordenschattesten onderzocht door de scores op de productieve en receptieve woordenschattesten te vergelijken met

de scores op de passieve woordenschattaak van de TAK. De twee woordenschattesten zijn ook met elkaar vergeleken. Het was tijdens dit onderzoek niet haalbaar om de productieve woordenschattest te vergelijken met een andere test die de productieve woordenschat meet. De productieve woordenschattest en de passieve woordenschattaak van de TAK meten niet precies hetzelfde construct. Toch werd er een positief verband verwacht tussen deze twee testen. De passieve woordenschattaak van de TAK correleert namelijk ook hoog met een productieve woordenschattaak (woordomschrijving) van de TAK (Verhoeven et al., 2006). Wel werd een sterker verband verwacht tussen de scores op de receptieve woordenschattest en de scores op de passieve woordenschattaak van de TAK aangezien deze testen precies hetzelfde construct meten. Dit wordt ook wel convergent bewijs genoemd, waarbij er hoge positieve correlaties gevonden worden tussen testen die hetzelfde construct meten en meer gematigde positieve correlaties tussen testen die gerelateerde constructen meten (Cohen et al., 2010).

Ook is de bruikbaarheid van de woordenschattesten bij het effectiviteitsonderzoek naar TPR onderzocht. Er is gekeken of de duur van de testen niet te belastend is voor de kinderen. Hiernaast is aan de hand van de moeilijkheidsgraden bepaald of de testen niet te makkelijk of te moeilijk zijn voor het effectiviteitsonderzoek naar TPR.

Resultatensectie

Beschrijvende statistiek

Tabel 2 laat zien dat de kinderen hogere scores hebben behaald op de receptieve woordenschattest dan op de productieve woordenschattest. Dit komt overeen met het feit dat de receptieve woordenschat van mensen altijd groter is dan de productieve woordenschat (Bossers, Kuiken & Vermeer, 2010). In tabel 2 valt ook af te lezen dat de kinderen het gemiddeld erg goed hebben gedaan op de passieve woordenschattaak van de TAK. De kinderen hebben op deze taak een gemiddelde score van 30.6 behaald. Volgens de handleiding van de TAK is dit een hoge score voor kinderen die thuis overwegend een andere taal dan het Nederlands spreken en een gemiddelde score voor kinderen die thuis overwegend Nederlands spreken. Aan de minimum en maximum

behaalde scores op de passieve woordenschattaak van de TAK valt af te lezen dat zowel kinderen met een zeer beperkte woordenschat als kinderen met een zeer uitgebreide woordenschat hebben deelgenomen aan het onderzoek. De laagste score op de passieve woordenschattaak van de TAK was 2 en de hoogste score was 56. Volgens de handleiding van de TAK is een score van 2, zelfs voor kinderen die thuis overwegend een andere taal dan het Nederlands spreken, een erg lage score. Een score van 56 op deze taak is zelfs een hoge score voor kinderen die thuis overwegend Nederlands spreken. Ook de standaardafwijkingen van de drie testen en de grafieken A1 t/m A3 in bijlage A laten zien dat er veel spreiding is in de scores op de drie testen. Door deze spreiding in de scores is het mogelijk om de betrouwbaarheid en de validiteit van de productieve en receptieve woordenschattesten vast te stellen (Goodwin et al., 2006). Nu zal allereerst de kwaliteit van de afzonderlijke items worden besproken. Vervolgens komen de betrouwbaarheid en de validiteit van de productieve en receptieve woordenschattesten aan bod.

Tabel 2

De Gemiddelde score, de Standaardafwijking en de Minimum en Maximum behaalde score (N=44) op de drie testen.

	Gemiddelde	Standaardafwijking	Minimum	Maximum
Productieve woordenschattest	8.5	4.58	0	17
Receptieve woordenschattest	22.6	5.51	7	30
Passieve woordenschattaak van de TAK	30.6	15.52	2	56

Itemanalyse productieve woordenschattest

In tabel 3 worden onder andere de item-restcorrelaties van de items van de productieve woordenschattest weergegeven en het effect dat het verwijderen van een item heeft op de betrouwbaarheid van de test. De items ‘omhoog’, ‘acrobaat’, ‘tussendoor’ en ‘in evenwicht’ zijn door alle kinderen niet of fout beantwoord en konden daardoor niet worden meegenomen in de betrouwbaarheidsanalyse. Om deze reden worden bij deze items geen waarden gerapporteerd voor ‘effect op alfa’. Aangezien alle kinderen deze vier items niet of fout hebben beantwoord, maken deze items geen onderscheid tussen kinderen met een beperkte woordenschat en kinderen met een uitgebreide woordenschat. Op dit moment hebben deze items weinig toegevoegde waarde en het behouden van deze items zal de test minder efficiënt maken. Er zal nog wel gekeken worden of het verwijderen van deze items invloed heeft op de validiteit van de productieve woordenschattest. Nu zal eerst de kwaliteit van de resterende items worden besproken.

In tabel 3 is te zien dat meerdere items een lagere item-restcorrelatie dan .20 hebben en enkele items zelfs negatief correleren met de rest van de test. Deze items meten mogelijk een ander construct dan de rest van de test, waardoor zij misschien beter verwijderd kunnen worden (Ebel, 1972). Na de verwijdering van de items ‘zonder’, ‘directeur’, ‘kassa’ en ‘onderdoor’ heeft het merendeel van de items een hogere item-restcorrelatie dan .30 en slechts 5 items een item-restcorrelatie tussen de .20 en .30. Volgens de normen van Ebel (1972) zullen deze 5 items nog wel verbeterd moeten worden. Wanneer de items ‘zonder’, ‘directeur’, ‘kassa’ en ‘onderdoor’ een negatieve invloed hebben op de betrouwbaarheid en de validiteit van de test is het beter om deze items te verwijderen. Aan de waarden van ‘effect van alfa’ in tabel 3 valt al af te lezen dat het verwijderen van deze items waarschijnlijk maar weinig impact zal hebben op de betrouwbaarheid van de productieve woordenschattest.

In tabel 3 worden ook de moeilijkheidsgraden van de items weergegeven. De kinderen vonden de items ‘zonder’, ‘directeur’, ‘kassa’ en ’onderdoor‘ erg moeilijk. Dit zou de lage item-restcorrelaties van deze items kunnen verklaren. Deze items zijn niet in staat geweest om de

kinderen met lage totaalscores te onderscheiden van kinderen met hoge totaalscores, doordat de items door alle kinderen erg moeilijk zijn gevonden. Ook de items ‘circus’, ‘elastiek’, ‘voorlezen’, ‘ronddraaien’, ‘hangen’, ‘stempelen’, ‘trommel slaan’, ‘gevaarlijk’, ‘hoepel’, ‘buigen’ en ‘goochelen’ hebben een lagere moeilijkheidsgraad dan .27 (Feldt, 1993). De hoogte van de item-restcorrelaties zijn voldoende en de items hebben volgens de waarden van ‘effect op alfa’ geen negatieve invloed op de betrouwbaarheid van de gehele test. Deze items kunnen dus bewaard blijven. De lage moeilijkheidsgraden hebben echter wel invloed op de bruikbaarheid van de test. Dit komt in de discussiesectie aan de orde.

Tabel 3

De Item-restcorrelatie, het Effect op Alfa, Moeilijkheidsgraad (p) en de Item-restcorrelatie na de Verwijdering van Items met een Lage Item-restcorrelatie voor alle Items van de Productieve Woordenschattest.

	Item-restcorrelatie	Effect op alfa[a]	Uiteindelijke item-restcorrelatie	*p*
1. fluit	.27	.84	.30	.64
2. circus	.46	.83	.42	.23
3. touw	.68	.82	.65	.77
4. aap	.55	.83	.55	.84
5. clown	.65	.82	.67	.61
6. elastiek	.23	.84	.24	.23
7. snor	.59	.82	.57	.57
8. mensen	.49	.83	.51	.56
9. voorlezen	.39	.83	.36	.25
10. ronddraaien	.38	.83	.39	.23
11. hangen	.26	.84	.28	.07
12. klimmen	.69	.82	.69	.73
13. stempelen	.30	.84	.30	.14
14. trommel slaan	.23	.84	.25	.02
15. zonder	-.06	.84	-	.07
16. omhoog[b]	0	-	-	0
17. gevaarlijk	.35	.84	.35	.18
18. directeur	-.01	.84	-	.02
19. microfoon	.30	.84	.32	.32
20. kassa	.06	.84	-	.09
21. piano	.29	.84	.27	.48
22. gitaar	.55	.83	.55	.68
23. hoepel	.33	.84	.36	.23
24. acrobaat[b]	0	-	-	0
25. masker	.43	.83	.44	.27
26. buigen	.47	.83	.50	.21
27. goochelen	.27	.84	.28	.02
28. tussendoor[b]	0	-	-	0
29. in evenwicht[b]	0	-	-	0
30. onderdoor	.09	.84	-	.09

[a] Cronbach's alfa van de test is .84. [b] De items 16, 24, 28 en 29 bevatten geen variantie en konden daardoor niet meegenomen worden in de betrouwbaarheidsanalyse.

Itemanalyse receptieve woordenschattest

Uit de itemanalyse van de receptieve woordenschattest komt naar voren dat alleen de items ‘voorlezen’ en ‘masker’ erg laag correleren met de rest van de test en het item ‘tussendoor’ negatief correleert met de rest van de test (zie tabel 4). Deze items kunnen beter verwijderd worden wanneer zij een negatieve invloed hebben op de betrouwbaarheid en de validiteit van de test. Ook bij de receptieve woordenschattest geldt dat het verwijderen van items waarschijnlijk maar weinig impact zal hebben op de betrouwbaarheid van de test. Dit valt te concluderen uit de waarden van ‘effect op alfa’ in tabel 4. Het is raadzaam om het item ‘tussendoor’ sowieso aan te passen, aangezien dit item negatief correleert met de rest van de test. Kinderen met een lage totaalscore hebben dit item vaker goed beantwoord dan kinderen met een hoge totaalscore. In bijlage D is te zien dat 50% van de kinderen gekozen heeft voor het alternatief ‘D’ in plaats van het juiste alternatief ‘C’. Dit item is mogelijk verwarrend, waardoor het nodig is om de alternatieven van dit item aan te passen. Na de verwijdering van de items ‘voorlezen’, ‘masker’ en ‘tussendoor’ hebben de overige items, op de items ‘microfoon’ en ‘acrobaat’ na, een hogere item-restcorrelatie dan .30. Volgens de normen van Ebel (1972) zullen de items ‘microfoon’ en ‘acrobaat’ nog wel verbeterd moeten worden.

Tabel 4 toont ook de moeilijkheidsgraden van de afzonderlijke items. Een groot deel van de items heeft een hogere moeilijkheidsgraad dan de bovengrens van .79 (Feldt, 1993). Tabel 4 toont ook de aangepaste moeilijkheidsgraden, waarbij gecorrigeerd is voor de gokkans (zie *p’*). Ook na aanpassing van de moeilijkheidsgraden hebben 10 items een moeilijkheidsgraad hoger dan .79. Toch hebben deze items volgens de waarden van ‘effect op alfa’ geen negatieve invloed op de betrouwbaarheid. Deze items hebben ook allemaal hogere item-restcorrelaties dan .30 en zijn dus in staat kinderen met een lage totaalscore te onderscheiden van kinderen met een hoge totaalscore. Het is om deze redenen niet nodig om deze items met hoge moeilijkheidsgraden aan te passen. Ondanks de hoge moeilijkheidsgraden is de receptieve woordenschattest waarschijnlijk goed te gebruiken bij het effectiviteitsonderzoek naar TPR. Dit zal in de discussie worden besproken.

Het item ‘acrobaat’ had juist een te lage moeilijkheidsgraad (Feldt, 1993). Dit zou de lagere item-restcorrelatie van dit item kunnen verklaren. In bijlage D is te zien dat de afleiders van dit item ongeveer even vaak zijn gekozen. De lage moeilijkheidsgraad is om deze reden waarschijnlijk niet het gevolg van onduidelijkheid van het item door een verwarrende afleider. De lage moeilijkheidsgraad kan eerder verklaard worden door de onbekendheid van dit woord.

Tabel 4

De Item-restcorrelatie, het Effect op Alfa, Moeilijkheidsgraad (p), Gecorrigeerde Moeilijkheidsgraad (p') en Item-restcorrelatie na de Verwijdering van Items met Lage Item-restcorrelaties voor alle Items van de Receptieve Woordenschattest.

	Item-restcorrelatie	Effect op Alfa[a]	Uiteindelijke item-restcorrelatie	*p*	*p'*
1. fluit	.41	.86	.41	.93	.91
2. circus	.64	.86	.67	.82	.76
3. touw	.46	.86	.46	.93	.91
4. aap	.54	.86	.53	.91	.88
5. clown	.59	.86	.56	.80	.73
6. elastiek	.56	.86	.59	.77	.69
7. snor	.50	.86	.48	.86	.81
8. mensen	.68	.86	.67	.82	.76
9. voorlezen	.05	.87	-	.82	.76
10. ronddraaien	.33	.86	.35	.89	.85
11. hangen	.46	.86	.50	.89	.85
12. klimmen	.37	.86	.40	.93	.91
13. stempelen	.63	.86	.65	.68	.57
14. trommel slaan	.54	.86	.56	.73	.64
15. zonder	.41	.86	.41	.71	.61
16. omhoog	.46	.86	.46	.93	.91
17. gevaarlijk	.48	.86	.47	.77	.69
18. directeur	.36	.86	.37	.57	.43
19. microfoon	.24	.87	.23	.68	.57
20. kassa	.38	.86	.39	.52	.36
21. piano	.33	.86	.34	.88	.84
22. gitaar	.39	.86	.40	.84	.79
23. hoepel	.45	.86	.46	.86	.81
24. acrobaat	.28	.87	.25	.43	.24
25. masker	.04	.87	-	.80	.73
26. buigen	.47	.86	.47	.57	.43
27. goochelen	.52	.86	.53	.59	.45
28. tussendoor	-.22	.88	-	.32	.09
29. in evenwicht	.50	.86	.51	.59	.45
30. onderdoor	.37	.86	.40	.80	.73

[a] Cronbach's alfa van de test is .87

Betrouwbaarheid Productieve en Receptieve woordenschattesten

Vervolgens is de betrouwbaarheid van de productieve woordenschattest onderzocht. De productieve woordenschattest had een Guttman's lambda2 van .86. Dit ligt ver boven de vereiste betrouwbaarheid van .60, die nodig is bij testen die gebruikt zullen worden bij onderzoek op groepsniveau (Evers et al., 2010). Bij de berekening van deze betrouwbaarheid zijn de items 'omhoog', 'acrobaat', 'tussendoor' en 'in evenwicht' niet meegenomen, omdat deze items geen variantie bevatten. Vervolgens is gekeken of de verwijdering van de items 'zonder', 'directeur', 'kassa' en 'onderdoor' een positieve invloed heeft op de betrouwbaarheid van de productieve woordenschattest. Na verwijdering van de items 'zonder', 'directeur', 'kassa' en 'onderdoor' had Guttman's lambda2 een waarde van .87. De verwijdering van deze items heeft dus maar weinig invloed op de betrouwbaarheid van de test.

Vervolgens is de betrouwbaarheid van de receptieve woordenschattest onderzocht. De receptieve woordenschattest had een Guttman's lambda2 van .88. Ook deze betrouwbaarheid ligt ver boven de vereiste betrouwbaarheid van .60 (Evers et al., 2010). Vervolgens is gekeken of de verwijdering van de items ' tussendoor', 'voorlezen' en 'masker' een positieve invloed heeft op de betrouwbaarheid van de receptieve woordenschattest. Na verwijdering van de items 'tussendoor', 'voorlezen' en 'masker' nam Guttman's lambda2 toe tot .90. Geconcludeerd kan worden dat ook zonder aanpassingen aan de productieve en receptieve woordenschattesten deze testen ruim voldoende betrouwbaar zijn voor het effectiviteitsonderzoek naar TPR. Dat het verwijderen van items met lagere item-restcorrelaties dan .20 maar weinig effect zou hebben op de betrouwbaarheid van de testen was al af te leiden uit de waarden van 'effect op alfa' in de tabellen 3 en 4.

Validiteit van de productieve en receptieve woordenschattesten

Nu zal de validiteit van de productieve en receptieve woordenschattesten besproken worden en zal gekeken worden of het verwijderen van items met lagere item-restcorrelaties dan .20 een positieve invloed heeft op de validiteit. Allereerst is de relatie tussen de scores op de

productieve woordenschattest en de scores op de passieve woordenschattaak van de TAK onderzocht aan de hand van Pearson's r. Er is een hoge positieve correlatie gevonden tussen de scores op de productieve woordenschattest en de scores op de passieve woordenschattaak van de TAK, r= .78. Volgens Cohen en collega's (2003) betekenen correlaties vanaf .50 dat er sterke verbanden bestaan. Er bestaat dus een zeer sterk verband tussen de scores op de productieve woordenschattest en de scores op de passieve woordenschattaak van de TAK. Vervolgens is gekeken of het verwijderen van de items 'omhoog', 'acrobaat', 'tussendoor' en 'in evenwicht' en de items 'zonder', 'directeur', 'kassa' en 'onderdoor' uit de productieve woordenschattest invloed heeft op de sterkte van dit verband. Na de verwijdering van deze items waren de scores op de productieve woordenschattest volgens de Kolmorogov-Smirnov test niet meer normaal verdeeld. Om deze reden is het verband tussen de productieve woordenschattest en de passieve woordenschattaak dit keer met behulp van Spearman's rho onderzocht (Field, 2009). Weer werd een sterk verband gevonden tussen de scores op de productieve woordenschattest en de scores op de passieve woordenschattaak van de TAK, r= .80.

Vervolgens is het verband tussen de scores op de receptieve woordenschattest en de scores op de passieve woordenschattaak van de TAK onderzocht. De verdeling van de scores op de receptieve woordenschattest was volgens de Kolmorogov-Smirnov test significant anders dan een normaal verdeling. Om deze reden is het verband tussen deze twee testen onderzocht met behulp van Spearmen's rho (Field, 2009). Er is een hoge positieve correlatie gevonden tussen de scores op de receptieve woordenschattest en de scores op de passieve woordenschattaak van de TAK, r= .71. Ook is gekeken of het verwijderen van de items 'tussendoor', 'voorlezen' en 'masker' uit de receptieve woordenschattest een positieve invloed heeft op de sterkte van dit verband. Na de verwijdering van deze items was de verdeling van de scores op de receptieve woordenschattest weer significant anders dan een normaal verdeling. Om deze reden is de relatie tussen de scores op receptieve woordenschattest en de scores op de passieve woordenschattaak van de TAK weer onderzocht met behulp van Spearman's rho (Field, 2009). Na de verwijdering van de items werd

weer een sterk positief verband gevonden tussen de scores op de receptieve woordenschattest en de scores op de passieve woordenschattaak van de TAK, r= .73.

Ten slotte is de relatie tussen de scores op de productieve woordenschattest en de scores op de receptieve woordenschattest onderzocht. Wegens de scheve verdeling van de scores op de receptieve woordenschattest is dit gedaan met behulp van Spearman's rho (Field, 2009). Tussen de scores op de twee woordenschattesten bestaat ook weer een sterk verband, r= .78. Hogere scores op de productieve woordenschattest gaan dus gepaard met hogere scores op de receptieve woordenschattest en vice versa . Vervolgens is weer gekeken of het verwijderen van de items 'omhoog', 'acrobaat', 'tussendoor', 'in evenwicht', 'zonder', 'directeur' en 'kassa' uit de productieve woordenschattest en de items 'tussendoor', 'voorlezen' en 'masker' uit de receptieve woordenschattest invloed heeft op de sterkte van dit verband. Weer werd er een sterk verband gevonden tussen de scores op de productieve woordenschattest en de scores op de receptieve woordenschattest met behulp van Spearman's rho, r= .77.

Het is opvallend dat er een hogere correlatie is gevonden tussen de scores op de productieve woordenschattest en de scores op de passieve woordenschattaak van de TAK dan tussen de scores op de receptieve woordenschattest en de scores op de passieve woordenschattaak van de TAK. Er werd juist een sterker verband verwacht met de receptieve woordenschattest. Testen die hetzelfde construct meten correleren namelijk sterker met elkaar dan testen die slechts gerelateerde constructen meten (Cohen et al., 2010). De correlatie tussen de scores op de productieve woordenschattest en de scores op de passieve woordenschattaak van de TAK (r = .78) is echter niet significant hoger dan de correlatie tussen de scores op de receptieve woordenschattest en de passieve woordenschattaak van de TAK (r= .71), t(41)= 1.12, p >.05. Ook na de verwijdering van items was dit verschil niet significant, t(41)= 1.15, p > .05.

De lagere correlatie tussen de scores op de receptieve woordenschattest en de scores op de passieve woordenschattaak van de TAK zou verklaard kunnen worden door de erg scheve verdeling van de scores op de receptieve woordenschattest en de erg hoge moeilijkheidsgraden van de items

van de receptieve woordenschattest. Correlaties tussen variabelen kunnen namelijk alleen de maximum waarde van 1.0 bereiken wanneer de vormen van de distributies van die variabelen gelijk zijn (Goodwin et al., 2006). En ook wanneer moeilijkheidsgraden veel afwijken van .50. zakt het plafond van de correlatie naar .80 (Goodwin et al., 2006). De receptieve woordenschattest is te makkelijk geweest voor deze steekproef, waardoor de moeilijkheidsgraden van de items hoog liggen en de scores scheef zijn verdeeld. Wanneer één van de verdelingen scheef is verdeeld wordt aangeraden om de Spearman's rank-order correlatie te berekenen in plaats van de Pearson product-moment correlatie (Bishara & Hittner, 2012). De scores worden dan eerst gerangschikt, voordat de Pearson product-moment correlatie wordt berekend. Het is op deze manier echter niet gelukt om een sterker verband te vinden tussen de scores op de receptieve woordenschattest en de scores op de passieve woordenschattaak van de TAK.

Een andere optie is om de verdeling van de scores eerst te transformeren, zodat een normaalverdeling ontstaat, voordat de Pearson correlatie wordt berekend. Op deze manier is het mogelijk om sterkere verbanden te vinden. Het is nodig om beide verdelingen te transformeren, zodat de testen met elkaar vergeleken kunnen worden. RIN transformaties zijn in staat om data normaal te verdelen ongeacht de vorm die de verdeling bij aanvang heeft. RIN transformaties en Spearman rank-order correlatie lijken conceptueel op elkaar. Het enige verschil is dat RIN transformaties na het rangschikken van de data en voor het berekenen van de Pearson correlatie de platte verdeling van de rangen transformeert in een normaalverdeling (Bishara et al., 2012). In bijlage E worden de verdelingen van de scores van de receptieve woordenschattest en de passieve woordenschattaak van de TAK voor en na het toepassen van een RIN transformatie weergegeven. Na het uitvoeren van een RIN transformatie op de verdelingen van de scores werd weer een Pearson correlatie van .72 gevonden. Het transformeren van de scores had dus geen invloed op de sterkte van het verband.

Discussie

De productieve en receptieve woordenschattesten zijn beide zeer betrouwbaar en valide in het meten van de woordenschat van vierjarige tweedetaalverwervers. Niet alle items voldeden aan de eis die werd gesteld aan de hoogte van de item-restcorrelatie. Het behouden van deze items heeft echter maar weinig invloed op de betrouwbaarheid en de validiteit van deze testen. De vraag is nu of deze items bewaard moeten blijven.

Uiteindelijk zullen deze testen gebruikt worden om vast te stellen of kinderen na een periode van drie weken met behulp van TPR meer doelwoorden van de woordenschattesten leren dan met behulp van de reguliere aanpak, zowel productief als receptief. Wanneer items uit de woordenschattesten worden verwijderd is het niet meer mogelijk om vast te stellen of de kinderen ook deze doelwoorden hebben kunnen verwerven. Bij het verwijderen van items zal dus informatie verloren gaan. Om deze reden heeft het de voorkeur om tijdens het effectiviteitsonderzoek naar TPR alle items bij de kinderen af te nemen. Het verwijderen van items met zeer lage item-restcorrelaties zal echter wel een positieve invloed hebben op het discriminerend vermogen en de efficiëntie van de testen. Een andere mogelijkheid is om tijdens het effectiviteitsonderzoek naar TPR alle items af te nemen en slechts de ‘goede’ items mee te nemen in de somscores, die gebruikt zullen worden om de effectiviteit van TPR vast te stellen.

Door alle items af te nemen is het ook mogelijk om de kwaliteit van alle items nogmaals te onderzoeken. Het is bijvoorbeeld interessant om tijdens het effectiviteitsonderzoek naar TPR het discriminerend vermogen van de items van de productieve woordenschattest met lage item-restcorrelaties nog een keer te onderzoeken. Zoals in de resultatensectie naar voren kwam, zijn de lage item-restcorrelaties van deze items waarschijnlijk het gevolg van de lage moeilijkheidsgraden. Het is echter te verwachten dat de moeilijkheidsgraden tijdens de nameting van het effectiviteitsonderzoek naar TPR hoger zullen liggen. Het is mogelijk dat deze items tijdens de nameting hierdoor een beter discriminerend vermogen hebben. Dit geldt ook voor de vier zeer

moeilijke vragen, die alle kinderen niet of fout beantwoord hebben. Het bewaren van deze zeer moeilijke items heeft echter ook invloed op de bruikbaarheid van de productieve woordenschattest.

Bruikbaarheid van de testen

Ten slotte stellen we de – niet onbelangrijke - vraag of de productieve en receptieve woordenschattesten bruikbaar zijn bij het onderzoek naar de effectiviteit van TPR. Aan het effectiviteitsonderzoek naar TPR zullen alleen kinderen met een beperkte woordenschat deelnemen. Om dit te bereiken zal tijdens het onderzoek naar TPR weer de passieve woordenschattaak van de TAK worden afgenomen. Alleen kinderen met een lagere score dan 30 op de passieve woordenschattaak van de TAK zullen deelnemen aan het onderzoek. De productieve woordenschattest is al erg moeilijk gevonden door de kinderen van het huidige onderzoek, terwijl deze kinderen het gemiddeld erg goed hebben gedaan op de passieve woordenschattaak van de TAK. De kinderen van het effectiviteitsonderzoek naar TPR zullen wegens hun beperkte woordenschat zeker moeite hebben met de productieve woordenschattest. De kinderen zullen tijdens de voormeting van het effectiviteitsonderzoek waarschijnlijk niet in staat zijn om de doelwoorden te produceren. Wanneer een kind tijdens de afname van de test geen van de woorden kan benoemen, zal dit tot veel stress bij het kind kunnen leiden en het kind demotiveren om nogmaals mee te werken aan deze test of zelfs aan de interventie. Hiermee zal rekening moeten worden gehouden. Volgens de ethische principes van de ‘American Psychological Association’ (2010) zal schade en stress die participanten mogelijk tijdens een onderzoek ervaren, zoveel mogelijk beperkt moeten worden. De kinderen zullen goed voorbereid moeten worden op de moeilijkheid van deze test. De kinderen zal duidelijk gemaakt moeten worden dat het niet erg is wanneer hij of zij een antwoord niet weet. Aangezien nu de moeilijkheidsgraden van de items bekend zijn, is het mogelijk om moeilijke items en makkelijke items goed te verspreiden, zodat moeilijke items niet te vaak op elkaar volgen.

Wegens de moeilijkheid van deze test is het misschien toch beter om de items met lage item-restcorrelaties en lage moeilijkheidsgraden te verwijderen. De kinderen van het huidige

onderzoek hadden al veel moeite met de moeilijkheid en de duur van de test. Door deze items te verwijderen zal de test makkelijker zijn en minder lang duren. Het verwijderen van deze moeilijke items zal de test een stuk minder belastend maken. Er zal echter goed gekeken moeten worden welke items precies verwijderd moeten worden. Wanneer alle items met lage item-restcorrelaties worden verwijderd zal de test namelijk geen doelwoorden meer bevatten uit het categorie 'overige woorden'. Ten slotte is het ook mogelijk om deze moeilijke items aan het einde van de test te plaatsen en alleen af te nemen onder kinderen die tijdens de afname van de test weinig stress ervaren. Er zal dan per kind gekeken moeten worden of afname van de gehele productieve woordenschattest mogelijk is.

In tegenstelling tot de productieve woordenschattest is de receptieve woordenschattest zeker bruikbaar bij kinderen met een beperkte woordenschat. De duur van deze test is niet te belastend. Het afnemen van deze test kost slechts enkele minuten. Tijdens het huidige onderzoek was er sprake van een plafondeffect. De test is tijdens het huidige onderzoek te makkelijk gebleken, waardoor een scheve verdeling van de scores is ontstaan en de moeilijkheidsgraden erg hoog liggen. Aangezien alleen kinderen met een beperkte woordenschat deelnemen aan het effectiviteitsonderzoek naar TPR, zal dit probleem zich waarschijnlijk niet meer voordoen. Naar verwachting zullen de moeilijkheidsgraden van de items van de receptieve woordenschattest bij deze doelgroep een stuk dichter bij het ideaal van .50 liggen (Feldt, 1993). Hierdoor zal er minder sprake zijn van een plafondeffect en is het mogelijk om progressie in de woordenschat van deze kinderen vast te stellen.

Aangezien tijdens het effectiviteitsonderzoek naar TPR weer de passieve woordenschattaak van de TAK wordt afgenomen is het mogelijk om nogmaals de begripsvaliditeit van de receptieve woordenschattest te onderzoeken. De kans op een scheve verdeling van de scores zal een stuk kleiner zijn bij deze specifieke doelgroep. Het is interessant om te onderzoeken of er een sterker verband gevonden wordt tussen de scores op de receptieve woordenschattest en de scores op de

passieve woordenschattaak van de TAK, wanneer beide verdelingen zonder transformaties normaal zijn verdeeld.

Beperkingen van het onderzoek

De kinderen hebben hogere scores behaald op de receptieve woordenschattest dan op de productieve woordenschattest. Dit komt overeen met het feit dat de receptieve woordenschat van mensen over het algemeen groter is dan de productieve woordenschat (Bossers et al., 2010). Tijdens het huidige onderzoek is de receptieve woordenschattest echter direct na de productieve woordenschattest afgenomen. Dit kan hebben bijgedragen aan de hoge scores op de receptieve woordenschattest. Het is mogelijk dat de kinderen tijdens de receptieve woordenschattest afbeeldingen hebben aangewezen die zij herkenden uit de productieve woordenschattest. Wanneer dit het geval is, wordt niet slechts de receptieve woordenschat gemeten. De test meet dan ook het geheugen van een kind. De receptieve woordenschattest meet dan niet wat het beoogt te meten. Één van de kinderen gaf tijdens de afname van de receptieve woordenschattest aan, de afbeeldingen van de doelwoorden te herkennen uit de productieve woordenschattest. Een mogelijke oplossing is het aanpassen van de afbeeldingen van de doelwoorden van de twee woordenschattesten, waardoor zij minder met elkaar overeenkomen. Een andere optie is om in het vervolg een korte pauze in te lassen tussen de afname van de productieve woordenschattest en de afname van de receptieve woordenschattest. Hierdoor zal de kans kleiner worden dat de kinderen de afbeeldingen herkennen uit de productieve woordenschattest.

Een andere beperking van het huidige onderzoek is de grootte van de steekproef. Bij een steekproefgrootte van 44 participanten bevatten de moeilijkheidsgraden nog een meetfout van ongeveer .07 (Dousma et al., 1997). De moeilijkheidsgraden die in tabel 3 en 4 worden gerapporteerd zijn dus niet de precieze moeilijkheidsgraden van de items. De moeilijkheidsgraden in tabel 3 en 4 geven natuurlijk wel een goede indicatie van de moeilijkheid van de items.

Conclusie

De productieve en receptieve woordenschattest zijn betrouwbaar en valide in het meten van de woordenschat van vierjarige tweedetaalverwervers. De receptieve woordenschattest zal zeker bruikbaar zijn bij het onderzoek naar de effectiviteit van TPR. Bij de productieve woordenschattest zal echter per kind gekeken moeten worden of de gehele afname van de test tijdens de voormeting van het effectiviteitsonderzoek naar TPR mogelijk is in verband met het stressniveau van het kind. Met behulp van de gevalideerde productieve en receptieve woordenschattesten is het nu mogelijk om onderzoek te doen naar het effect van TPR op zowel de productieve als de receptieve woordenschat van deze jonge tweedetaalverwervers. Wanneer TPR effectief blijkt, zal deze nieuwe aanpak ervoor kunnen zorgen dat jonge tweedetaalverwervers beter mee kunnen doen in de klas en hierdoor meer van het onderwijs kunnen profiteren. TPR zal één van de eerste aanpakken in Nederland zijn, die naast een goede theoretische onderbouwing ook empirische onderzocht is met behulp van gevalideerde meetinstrumenten.

Literatuur

American Psychological Association (2002). *Ethical principles of psychologists and code of conduct*. Verkregen van: http://www.apa.org/ethics/code/index.aspx.

Asher, J. J. (1988). *Learning another language through actions*: *The complete teacher's guidebook.* Los Gatos, California: Sky Oaks Productions Inc.

Bacchini, S., Boland, T., Hulsbeek, M., Pot, H. & Smits, M. (2005). *Duizend-en-een woorden. De allereerste Nederlandse woorden van anderstalige peuters en kleuters.* Enschede: SLO. Verkregen van: http://www.slo.nl/downloads/archief/Duizend-en-een-woorden.pdf.

Bacchini, S. (2012). *Eerste hulp bij tweede taal: Experimentele studies naar woordenschatdidactiek voor jonge tweede-taalverwervers* (dissertatie Universiteit van Amterdam). Amsterdam: UvA. Verkregen van: http://dare.uva.nl/record/432382.

Bender, W. (1994). Statistiek en meten: Wat moet je daarover weten? *Bulletin Medisch Onderwijs, 13*, 37-43. Verkregen van: http://www.nvmo.nl/nvmo_tijdschrift/archief_bulletin _med isch_onderwijs_19821999_.

Bishara, A. J. & Hittner, J. B. (2012). Testing the significance of a correlation with non-normal data: Comparison of Pearson, Spearman, transformation and resampling approaches. *Psychological Methods, 17*(3), 399-417. doi:10.1037/a0028087.

Bossers, B. H., Kuiken, F. & Vermeer, A. (2010). *Handboek Nederlands als tweede taal in het volwassenonderwijs.* Bussum: Coutinho.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates. Verkregen van: http://books.google.nl/books?hl=nl&lr=&id=gkalyqTMXNEC&oi=fnd&pg =PP1&dq=cohen+Applied+multiple+regression/correlation.

Cohen, R. J. & Swerdlik, M. E. (2010). *Psychological testing and assessment: An introduction to test & measurement* (pp. 139-147). New York: McGraw-Hill.

Cronbach, L. J. (1951). Coëfficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334. doi:10.1007/BF02310555.

Dekker, J. & Gelauff-Hanzon, C. (2000). *Keuzegids VVE: Programma's, deskundigheidsbevordering en instrumenten op het gebied van de voor- en vroegschoolse educatie.* Den Haag/Utrecht: Makelaar VVE.

Dousma, T., Horsten, A. & Brants, J. (1997). *Tentamineren.* (3e druk). Groningen: Wolters-Noordhoff.

Driessen, G. Doesborgh, J. (2003). *Voor- en vroegschoolse educatie en cognitieve en niet-cognitieve competenties van jonge kinderen.* Nijmegen: ITS. Verkregen van: http://scholar.google.com.proxy.library.uu.nl/scholar?hl=en&q=Voor+en+vroegschoolse+educatie+en+cognitieve+en+niet-cognitieve+competenties+van+jonge+kinderen.

Driessen, G. (2012). *Variatie in voor- en vroegschoolse educatie: Een onderzoek naar de uiteenlopende wijzen waarop in gemeenten vormgegeven wordt aan VVE*. Nijmegen: ITS, Radboud Universiteit Nijmegen. Verkregen van: http://www.vveplatform.info/bibliotheek/.

Ebel, R. L. (1972). *Essentials of educational measurement.* (2 ed.). Englewood Cliffs: Prentice-Hall Inc. Verkregen van: http://psycnet.apa.org/psycinfo/1973-22100-000.

Evers, A., Lucassen, W., Meijer, R. & Sijtsma, K. (2010). *COTAN beoordelingssysteem voor de kwaliteit van tests.* Zaandijk: Heijnis & Schipper. Verkregen van: http://scholar.google.nl/scholar?q=COTAN+beoordelingssysteem+voor+de+kwaliteit+van+tests.

Feldt, L. S. (1993). The relationship between the distribution of item difficulties and test reliability. *Applied Measurement in Education, 6*(1), 37-48. doi:10.1207/s15324818ame0601_3

Field, A. (2009). *Discovering statistics using SPSS*. London: SAGE Publications.

Gijsberts, M., Huijnk, W. & Dagevos, J. (2012). *Jaarrapport integratie 2011.* Den Haag: Sociaal en Cultureel Planbureau. Verkregen van: http://www.scp.nl/Publicaties/Alle_publicaties/Publicaties_2012/Jaarrapport_integratie_2011.

Gillis, S. & Schaerlaekens, A. (2000). *Kindertaalverwerving: Een handboek voor het Nederlands.* Groningen: Martinus Nijhoff.

Goodwin, C. J. (2008). *Research in psychology: methods and design* (pp. 124-128). Crawfordsville: R.R. Donnelly & Sons.

Goodwin, L. D. & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r. *The Journal of Experimental Education, 74*(3), 251-266. doi:10.3200/JEXE.74.3.249-266.

Guttman, L. (1945). A basis for analysing test-retest reliability. *Psychometrika, 10,* 255-282. doi: 10.1007/BF02288892.

Haan, A. de, Leseman, P. & Elbers, E. (2011). *Pilot gemengde groepen 2007-2010. Onderzoeksrapportage oktober 2011.* Utrecht: Universiteit van Utrecht, Orthopedagogiek. Verkregen van: http://www.utrecht.nl/utrechts-onderzoek-en-cijfers/onderzoeksrapporten /onderwijs-studeren/.

Hartgers, M. *(2012). Jaarrapport integratie: Onderwijs.* Den Haag: Centraal Bureau van de Statistiek. Verkregen van: http://www.cbs.nl/nl-NL/menu/themas/dossiers/allochtonen/publicaties /publicaties/archief/2012/2012-b61-pub.htm.

Helms-Lorenz, M. & Jong-Heeringa, J. L. de (2006). *Ontwikkelingsgericht woordenschatonderwijs: De implementatie en effecten bij Marokkaanse kleuters.* Groningen: GION. Verkregen van: http://gion.gmw.eldoc.ub.rug.nl/root/2006/ontwikkelingsgericht/.

Henson, R.K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coëfficient alpha. *Measurement and Evaluation in Counseling and Development, 34*(4), 177-179. Verkregen van: http://web.a.ebscohost.com.proxy.library.uu.nl/ehost/detail?sid 71276ea0-99e7-478d-bb9c-411913219c91%40sessionmgr4005&vid=1&hid=4212&bdata =JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#db=aph&AN=5542030.

Jong-Heeringa, J. L. de & Bosker, R. J. (2006). *Spelenderwijs in Huizen – implementatie en effecten van Spelenderwijs*. Groningen: GION. Verkregen van: http://scholar.google.com.proxy.library .uu.nl/scholar?Hl=en&q=implementatie+en+effecten+van+Spelenderwijs.

Kuiken, F. & Droge, S. (2010). *Woordenlijst Amsterdamse Kinderen: Deelproject 'woordenschat' taalproject Uva-RUG.* Universiteit van Amsterdam. Verkregen van: http://scholar.google.com .proxy.library.uu.nl/scholar?hl=en&q=kuiken+en+droge%2C+WAK.

Ledoux, G. & Veen, A. (2009). *Beleidsdoorlichting onderwijsachterstandenbeleid: periode 2002-2008.* Amsterdam: SCO-Kohnstamm Instituut. Verkregen van: http://scholar.google.com.proxy. library.uu.nl/scholar?q=Beleidsdoorlichting+onderwijsachterstandenbeleid.

Meij, H., Mutsaers, K. & Pennings, T. (2009). *Effectiviteit van voor- en vroegschoolse programma's in Nederland.* Utrecht: Nederlands Jeugdinstituut. Verkregen van: http://www.nji.nl/nl/Public aties/Effectiviteit-van-voor--en-vroegschoolse-programmas-inNederland.

Mulder, F., Timman, Y. & Verhallen, S. (2009). *Handreiking bij de Basiswoordenlijst Amsterdamse Kleuters (BAK).* Verkregen van: http://scholar.google.com.proxy.library.uu.nl/scholar ?hl=en&q=basiswoordenlijst+amsterdamse+kleuters.

Nap-Kolhoff, E., Schilt-Mol, T. van, Simons, M., Sontag, L., Steensel, R. van & Vallen, T. (2008). *VVE onder de loep. Een studie naar de uitvoering en effectiviteit van voor- en vroegschoolse educatieve programma's.* Tilburg: IVA. Verkregen van: http://www.vveplatform.info /bibliotheek/.

Pennings, T. (2006). *Databank effectieve jeugdinterventies: beschrijving 'KO-totaal'*. Utrecht: Nederlands Jeugdinstituut. Verkregen van: http://www.nji.nl/nl/Kennis/Databanken/ Databank- Effectieve-Jeugdinterventies/Interventies-VVE-catalogus.

Pot, H. (2006). *Een lawine van woorden: Taal leren door (absolute) beginners met Puk & Ko en Piramide in Rotterdam*. InHolland. Verkregen van: http://www.inholland.nl/onderzoek/ Lectoraten/Dynamiek+van+de+stad/Publicaties/Atrikelen+en+opinie/.

Pot, H. (2012). T_2PR bij jonge absolute beginners, een omweg waard? *Tijdschrift Taal*, *3*(5), 32-38. Verkregen van: http://www.inholland.nl/onderzoek/Lectoraten/Ontwikkelingsgericht +onderwijs/Publicaties/Publicaties+2012.htm.

Sarid, M. & Breznitz, Z. (1997). Developmental aspects of sustained attention among 2- to 6-year-old children. *International Journal of Behavioral Development, 21*(2), 303-312. doi:10.1080 /016502597384884.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107-120. doi:10.1007/s11336-008-9101-0.

Sylva, K., Melhuish, E. C., Sammons, P., Siraj-Blatchford, I. & Taggart, B. (2004). *The Effective Provision of Pre-school Education (EPPE) project: Final report: A longitudinal study funded by the DfES 1997-2004.* London: DfES / Institute of Education. Verkregen van: http://eprints .ioe.ac.uk/id/eprint/5309.

Tesser, P. & Iedema, J. (2001). *Rapportage minderheden 2001: vorderingen op school*. Den Haag: Sociaal & Cultureel Planbureau. Verkregen van: http://www.scp.nl/Publicaties/Alle_publi caties/Publicaties_2001/Rapportage_minderheden_2001_Deel_I.

Veen, A., Fukkink, R. & Roeleveld, J. (2006). *Evaluatie van Startblokken en Basisontwikkeling: Implementatie en effecten*. Amsterdam: SCO-Kohnstamm Instituut. Verkregen van: http:// scholar.google.com.proxy.library.uu.nl/scholar?q=Evaluatie+van+Startblokken+en+Basisont wikkeling.

Veen, A., Roeleveld, J. & Leseman, P. (2000). *Evaluatie van Kaleidoscoop en Piramide eindrapportage 2000.* Amsterdam: SCO-Kohnstamm Instituut. Verkregen van: http://scholar.google.com .proxy.library.uu.nl/scholar?q=Evaluatie+van+Kaleidoscoop+en+Piramide.

Verhoeven, L., Leeuwen, J. van & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading, 15*(1), 8-25. doi:10.1080/ 10888438.2011.536125.

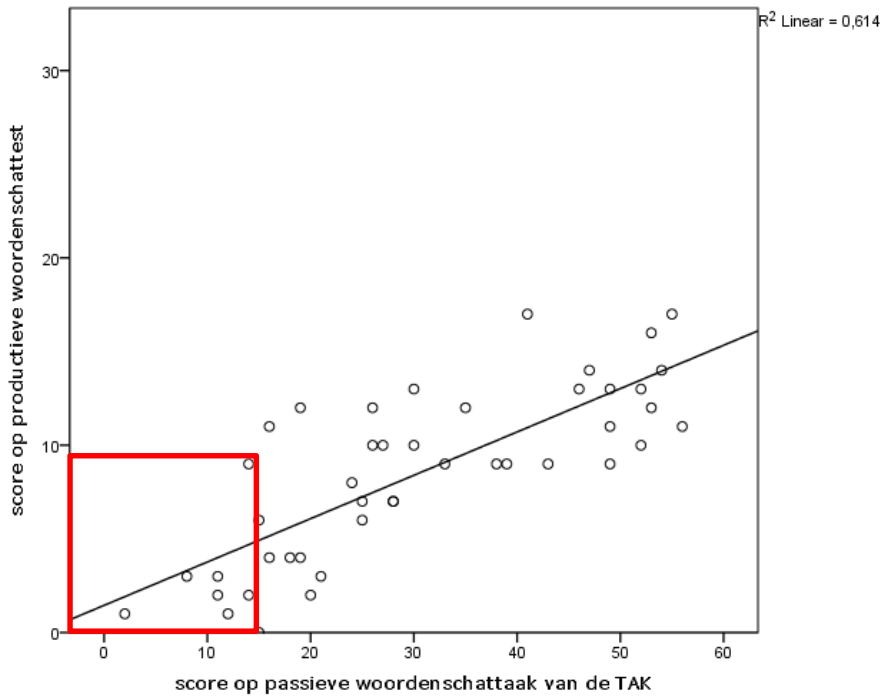
Verhoeven, L. & Vermeer, A. (1992). Woordenschat van leerlingen in het basis- en MLK-onderwijs. *Tijdschrift voor Onderwijskunde en Opvoedkunde*,*69*, 218-234. Verkregen van: http://objects .library.uu.nl.proxy.library.uu.nl/reader/resolver.php?obj=000739351.

Verhoeven, L. & Vermeer, A. *(2006). Verantwoording Taaltoets Alle Kinderen (TAK)*. Arnhem: Cito B.V.. Verkregen van: http://zoeken.kennisnet.nl/?q=verantwoording+taaltoest+alle+kinderen.

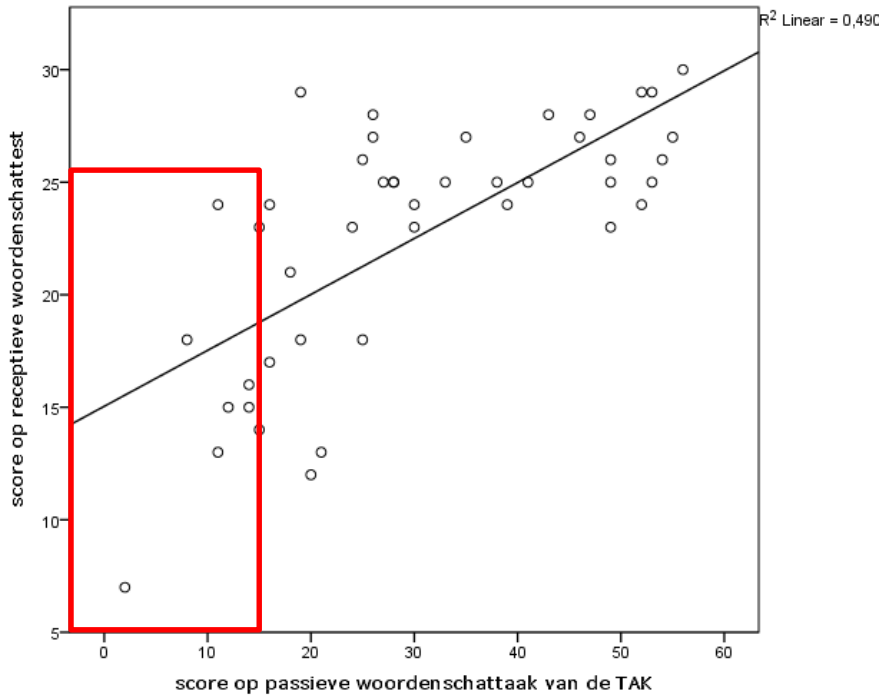
Bijlage A

De relaties tussen de scores van de drie testen.

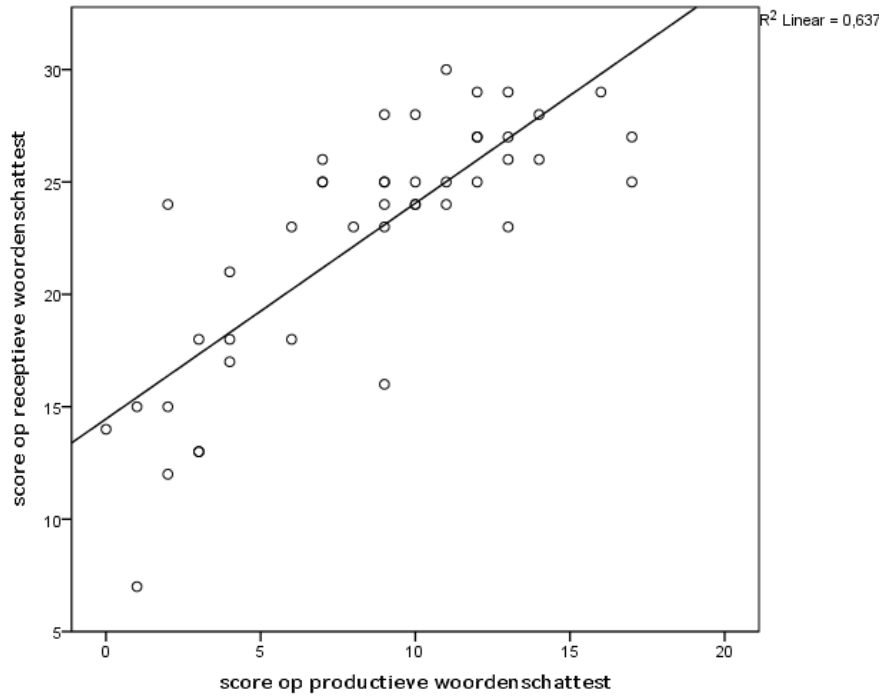
In de figuren A1 t/m A3 worden de verbanden tussen de productieve woordenschattest, de receptieve woordenschattest en de passieve woordenschattaak van de TAK weergegeven. Met behulp van de rode omkaderingen in de figuren A1 en A2 wordt geïllustreerd wat het effect van een beperkt bereik is op de sterkte van verbanden.



Figuur A1. Verband tussen de scores op de productieve woordenschattest en de scores op de passieve woordenschattaak van de TAK.



Figuur A2. Verband tussen de scores op de receptieve woordenschattest en de scores op de passieve woordenschattaak van de TAK.



Figuur A3. Verband tussen de scores op de productieve woordenschattest en de scores op de receptieve woordenschattest.

Bijlage B

De informatiebrief voor de scholen.

Heidelberglaan 1, De Uithof, Utrecht

Afzender: OWP, Postbus 80140, 3508 TC, Utrecht

Faculteit Sociale Wetenschappen

Departement Psychologie

Ontwikkelingspsychologie

Faxnummer
(030) 253 47 18
Telefoon
(030) 253 27 51
E-mail
D.M.P.deHaan@uu.nl
Blad
Blad 43 van 50

Datum:

Betreft: onderzoek naar de kwaliteit van een woordenschattoets, dat is gebaseerd op het vroegschoolse educatieprogramma 'Ik & Ko'.

Geachte heer/ mevrouw,

Vanaf september dit jaar start een promotieonderzoek van Hanneke Pot naar de effectiviteit van een methodiek voor taalonderwijs aan jonge kinderen die nog weinig Nederlands beheersen als ze naar school komen. Het onderzoek sluit aan bij het vroegschoolse educatieprogramma 'Ik & Ko'. Het is van belang om vooraf aan dit onderzoek een goede woordenschattoets te ontwikkelen. Het is voor dit deel van het onderzoek niet nodig dat uw school gebruik maakt van het VVE-programma 'Ik & Ko'. In dit stadium van het onderzoek willen we de betrouwbaarheid en validiteit van de toets onderzoeken. Hiervoor zijn wij op zoek naar vierjarige kinderen, die thuis een andere taal dan het Nederlands spreken. Wij hopen dat uw school interesse heeft om mee te werken met het onderzoek naar de kwaliteit van deze woordenschattoets.

De kinderen zullen één voor één uit de klas gehaald worden voor het afnemen van de toets door Stephanie Hoeber, masterstudent aan de Universiteit van Utrecht. Een kind zal in totaal twee keer meegenomen worden, omdat we drie onderdelen willen toetsen: de actieve en de passieve woordenschattoets voor het programma en

de passieve woordenschattest van de TAK (Taaltoets Alle Kinderen). In totaal zal een kind maximaal 15 minuten bezig zijn met de toetsen. Voor dit onderzoek hebben wij, naast het geslacht, de leeftijd van het kind en de gesproken thuistaal, geen persoonlijke informatie nodig. Het liefst nemen we de toetsen af op maandag of dinsdag wanneer de kinderen zich nog goed kunnen concentreren. Het is de bedoeling dat de toetsen in maart of april worden afgenomen.

Als u verder vragen heeft, kunt u contact opnemen met Stephanie Hoeber of Dorian de Haan (begeleider vanuit de Universiteit Utrecht). Graag nemen wij binnenkort telefonisch contact met u op om te vernemen of u wil meewerken aan dit onderzoek.

Met vriendelijke groet,

Stephanie Hoeber, Masterstudent Kind- en Jeugdpsychologie, Universiteit van Utrecht

Dr. Dorian. de Haan, docent Kind- en Jeugdpsychologie, Universiteit van Utrecht

Hanneke Pot, docent PABO Hogeschool Inholland, Rotterdam, promovenda.

Bijlage C

Toestemmingsformulier voor de ouders.

Heidelberglaan 1, De Uithof, Utrecht

Afzender: OWP, Postbus 80140, 3508 TC, Utrecht

Faculteit Sociale Wetenschappen

Departement Psychologie

Ontwikkelingspsychologie

Faxnummer
(030) 253 47 18
Telefoon
(030) 253 27 51
E-mail
D.M.P.deHaan@uu.nl

Datum:

Blad
Blad 45 van 50

Betreft: de deelname van uw kind aan een onderzoek.

Geachte ouder of verzorger,

Mijn naam is Stephanie Hoeber en ik ben masterstudent Kind- en Jeugdpsychologie aan de Universiteit van Utrecht. Ik doe op dit moment onderzoek naar de kwaliteit van een woordenschattest. Hiervoor is het nodig dat ik drie kleine woordenschattesten afneem onder kinderen van vier jaar. Hierbij vraag ik u om uw toestemming voor de deelname van uw kind.

Ik zal drie woordenschattesten bij de kinderen afnemen. Twee testen die voor het huidige onderzoek zijn ontwikkeld en één al bestaande test, de ‘Taaltoets Alle Kinderen’. Bij elkaar duren de testen ongeveer een kwartier. De testen zullen niet in één keer afgenomen worden aangezien dat te veel vraagt van een kind van vier. Bij twee van de testen zal uw kind gevraagd worden om uit vier plaatjes het juiste plaatje aan te wijzen en bij de derde test worden plaatjes getoond die uw kind vervolgens moet benoemen. Alleen ik en mijn begeleidsters zullen de antwoorden van uw kind kunnen inzien. Achteraf zal niet meer te zien zijn welke antwoorden uw kind heeft gegeven.

Deelname aan deze studie is geheel vrijwillig. Uw beslissing over de deelname van uw kind zal geen verdere gevolgen hebben. Zelfs wanneer u toestemming geeft, mag u kind nog weigeren en uw kind mag op elk moment stoppen met de test.

Wanneer u vragen heeft of meer informatie wil ontvangen kunt u mij bellen (……..) of mailen (……..). Met onderstaande strookje kunt u aangeven of uw bezwaar heeft tegen de deelname van uw kind.

Met vriendelijke groet,

Stephanie Hoeber, Masterstudent Kind- en Jeugdpsychologie Universiteit van Utrecht

Dr. Dorian. de Haan, docent Kind- en Jeugdpsychologie, Universiteit van Utrecht

Hanneke Pot, docent PABO Hogeschool Inholland, Rotterdam, promovenda.

.....

Hierbij geef ik aan dat ik niet wil dat mijn kind
(naam)……...
deelneemt aan het onderzoek naar de kwaliteit van een woordenschattest.

Naam ouder of verzorger :……

Handtekening ouder of verzorger : ……

Datum : ……

Bijlage D

Het percentage kinderen die een bepaalde antwoordmogelijkheid van een item van de receptieve woordenschattest heeft gekozen.

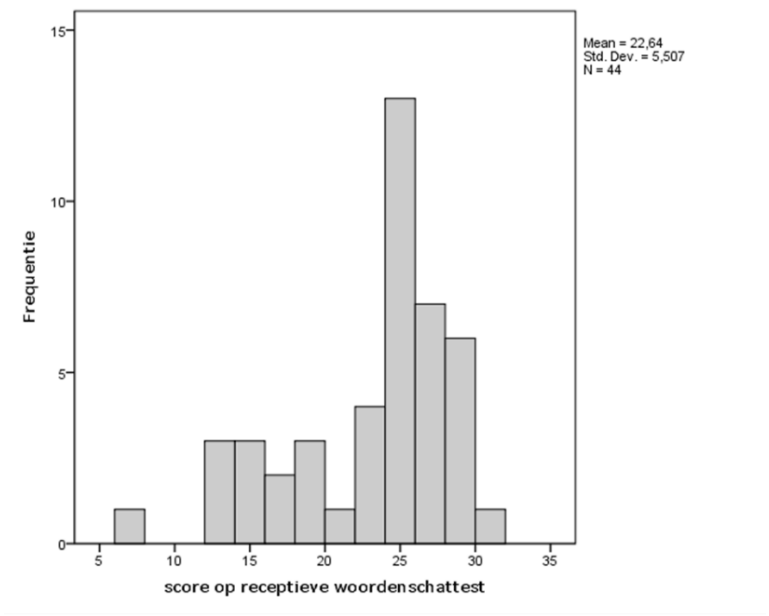
	A	B	C	D
1. fluit	93.2	4.5	2.3	-
2. circus	11.4	81.8	4.5	2.3
3. touw	-	2.3	93.2	4.5
4. aap	90.9	6.8	2.3	-
5. clown	4.5	15.9	-	79.5
6. elastiek	-	77.3	15.9	6.8
7. snor	9.1	2.3	86.4	2.3
8. mensen	2.3	9.1	6.8	81.8
9. voorlezen	-	81.8	15.9	2.3
10. ronddraaien	2.3	9.1	88.6	-
11. hangen	88.6	4.5	2.3	4.5
12. klimmen	93.2	-	6.8	-
13. stempelen	13.6	13.6	4.5	68.2
14. trommel slaan	2.3	72.7	2.3	22.7
15. zonder	27.3	-	70.5	2.3
16. omhoog	2.3	93.2	4.5	-
17. gevaarlijk	11.4	9.1	2.3	77.3
18. directeur	18.2	56.8	15.9	9.1
19. microfoon	11.4	68.2	2.3	6.8
20. kassa	20.5	15.9	11.4	52.3
21. piano	2.3	88.6	2.3	6.8
22. gitaar	84.1	6.8	-	9.1
23. hoepel	4.5	6.8	86.4	2.3
24. acrobaat	15.9	43.2	25.0	15.9
25. masker	13.6	2.3	4.5	79.5
26. buigen	56.8	6.8	25.0	11.4
27. goochelen	15.9	20.5	4.5	59.1
28. tussendoor	9.1	9.1	31.8	50.0
29. in evenwicht	6.8	20.5	59.1	13.6
30. onderdoor	79.5	4.5	11.4	4.5

Noot: de juiste antwoordmogelijkheden zijn gecentreerd.

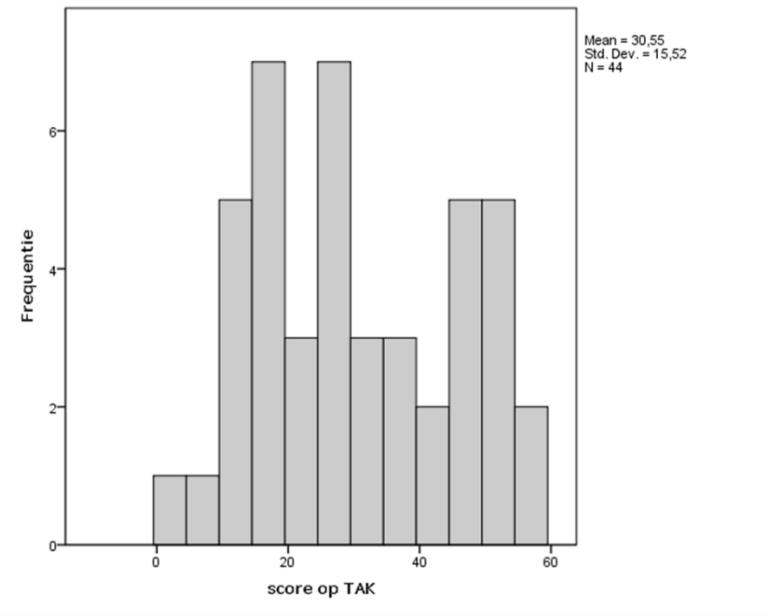
Bijlage E

Verdelingen van de scores op de receptieve woordenschattest en de scores op de passieve woordenschattaak van de TAK.

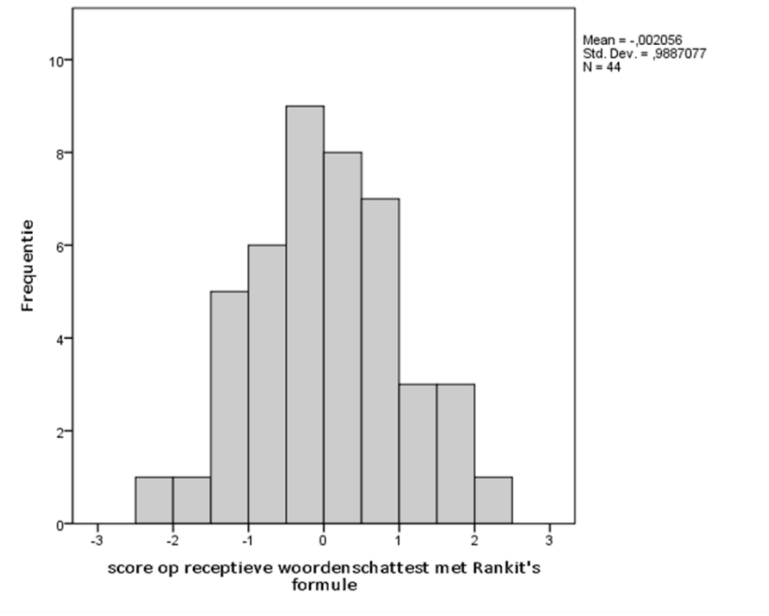
In de figuren E1 t/m E4 worden de verdelingen van de scores op de receptieve woordenschattest en de passieve woordenschattaak van de TAK voor en na het uitvoeren van een RIN transformatie weergegeven.



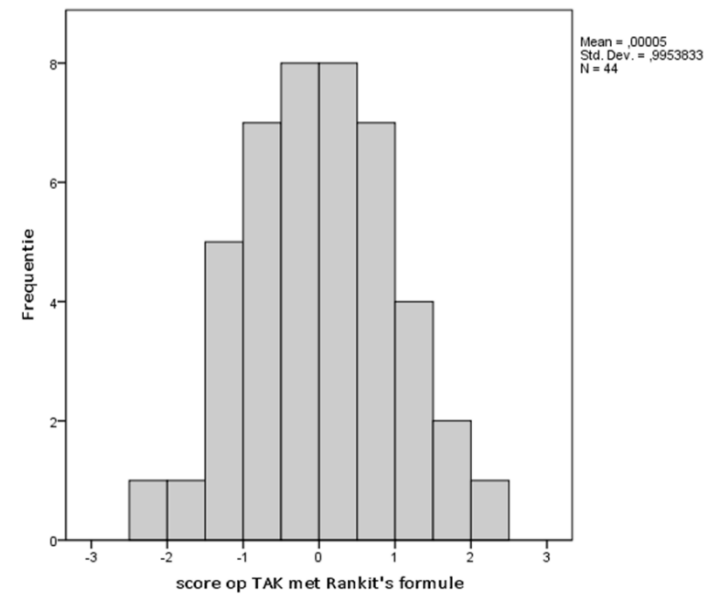
Figuur E1. Verdeling van de scores op de receptieve woordenschattest.



Figuur E2. Verdeling van de scores op de passieve woordenschattaak van de TAK.



Figuur E3. Verdeling van de scores op de receptieve woordenschattest na een RIN transformatie.



Figuur E4. Verdeling van de scores op de passieve woordenschattaak van de TAK na een RIN transformatie.